

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-105191

(43)Date of publication of application : 24.04.1998

(51)Int.Cl.

G10L 3/02

G10L 3/00

G10L 9/00

(21)Application number : 08-259967

(71)Applicant : TOSHIBA CORP

(22)Date of filing : 30.09.1996

(72)Inventor : KANAZAWA HIROSHI

ISAKA TAKEHIKO

NAGATA HITOSHI

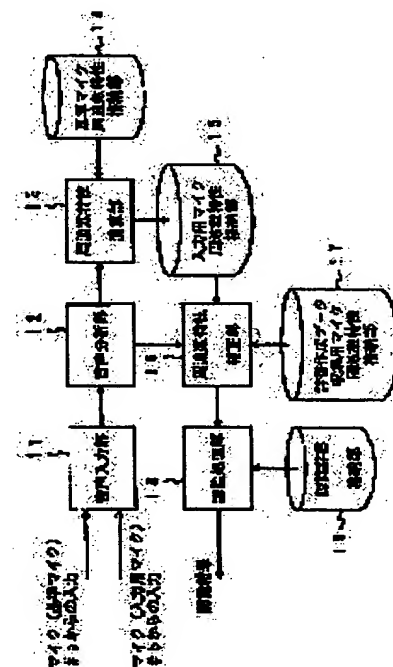
TSUBOI HIROYUKI

(54) SPEECH RECOGNITION DEVICE AND MICROPHONE FREQUENCY CHARACTERISTIC CONVERTING METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To accurately correct a difference in frequency characteristics between an input microphone and a microphone for dictionary generation without placing any burden on the user.

SOLUTION: A user's spoken speech is inputted by a speech input part 11 through a reference microphone #a and the input microphone #b at the same time to gather speech data for each microphone, and a speech analysis part 12 takes a frequency analysis of both the speech data; and a frequency characteristic arithmetic part 13 obtains the frequency characteristics of the input microphone #b from the difference between both analytic results and the frequency characteristics of the reference microphone #a in a storage part 14 and stores them in a storage part 15. For recognition, a frequency characteristic correction part 16 corrects frequency parameters for the input speech form the input microphone #b by using the differences between the frequency characteristics of a microphone for dictionary generation in a storage part 17 and the frequency characteristics of the input microphone #b in the storage part 15 to perform conversion into the frequency characteristics of the microphone for dictionary generation, so that they are used for recognition processing by a recognition processing part 18.



LEGAL STATUS

[Date of request for examination] 08.09.2000

[Date of sending the examiner's decision of rejection] 09.03.2004

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision]

of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-105191

(43) 公開日 平成10年(1998) 4月24日

(51) Int.Cl.⁶

G 1 0 L 3/02

識別記号

3 0 1

F I

G 1 0 L 3/02

3 0 1 A

3 0 1 E

3/00

5 1 1

3/00

5 1 1

9/00

9/00

F

審査請求 未請求 請求項の数 8 O L (全 17 頁)

(21) 出願番号

特願平8-259967

(22) 出願日

平成 8 年 (1996) 9 月 30 日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 金澤 博史

大阪府大阪市北区大淀中 1 丁目 1 番 30 号

株式会社東芝関西支社内

(72) 発明者 井阪 岳彦

大阪府大阪市北区大淀中 1 丁目 1 番 30 号

株式会社東芝関西支社内

(72) 発明者 永田 仁史

大阪府大阪市北区大淀中 1 丁目 1 番 30 号

株式会社東芝関西支社内

(74) 代理人 弁理士 鈴江 武彦 (外 6 名)

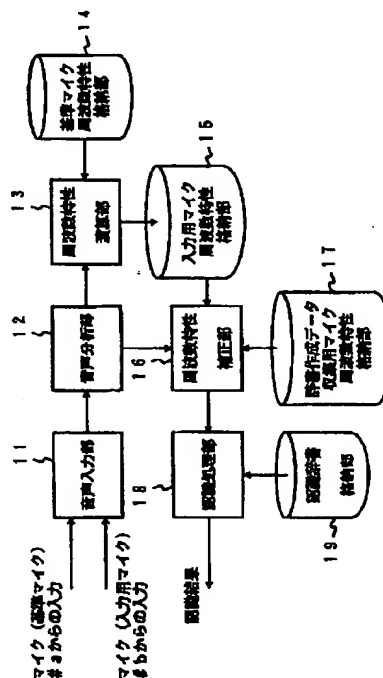
最終頁に続く

(54) 【発明の名称】 音声認識装置及びマイクロホン周波数特性変換方法

(57) 【要約】

【課題】 入力用マイクロホンと辞書作成用マイクロホンとの周波数特性の差をユーザに負担をかけずに正確に補正できるようにする。

【解決手段】 基準マイク # a と入力用マイク # b の両方を用いてユーザ発声の音声を音声入力部 11 により同時に入力することで各マイク毎に音声データを収集し、両音声データを音声分析部 12 にて周波数分析し、両分析結果の差分と格納部 14 内の基準マイク # a の周波数特性とから入力用マイク # b の周波数特性を周波数特性演算部 13 にて求め、格納部 15 に格納する。認識時には、入力用マイク # b からの入力音声に対して、格納部 17 内の辞書作成用マイクロホンの周波数特性と格納部 15 内の入力用マイク # b の周波数特性との差分を用いて周波数特性補正部 16 にて周波数パラメータの補正を行なうことで、辞書作成用マイクロホンの周波数特性に変換し、認識処理部 18 での認識処理に供する。



【特許請求の範囲】

【請求項1】 周波数特性が既知の基準マイクロホンと周波数特性が未知の入力用マイクロホンを用いて、ユーザが発声する音声を前記基準マイクロホンと前記入力用マイクロホンとで同期がとれるように入力する入力手段と、

前記入力手段により前記基準マイクロホン及び前記入力用マイクロホンから入力された音声を周波数分析する分析手段と、

前記分析手段による前記基準マイクロホン及び前記入力用マイクロホンからの入力音声の周波数分析結果の周波数成分の差分のデータを求め、その差分のデータと前記基準マイクロホンの周波数特性とから、前記入力用マイクロホンの周波数特性を求める周波数特性演算手段と、前記入力用マイクロホンから入力された認識対象音声に対して、予め取得されている辞書作成用マイクロホンの周波数特性と前記周波数特性演算手段により求められた入力用マイクロホンの周波数特性との周波数成分の差分のデータを用いて周波数パラメータの補正を行ない、前記辞書作成用マイクロホンの周波数特性に変換する周波数特性補正手段と、

前記周波数特性補正手段により変換された周波数パラメータからなる認識用音声データを認識辞書と照合することにより認識処理し、認識結果を出力する認識処理手段とを具備することを特徴とする音声認識装置。

【請求項2】 予め記憶してある音波形を用いて音を出力する音出力手段と、

周波数特性が既知の基準マイクロホンと周波数特性が未知の入力用マイクロホンを用いて、前記音出力手段から出力された音を前記基準マイクロホンと前記入力用マイクロホンとで同期がとれるように入力する入力手段と、前記入力手段により前記基準マイクロホン及び前記入力用マイクロホンから入力された音を周波数分析する分析手段と、

前記分析手段による前記基準マイクロホン及び前記入力用マイクロホンからの入力音の周波数分析結果の周波数成分の差分のデータを求め、その差分のデータと前記基準マイクロホンの周波数特性とから、前記入力用マイクロホンの周波数特性を求める周波数特性演算手段と、前記入力用マイクロホンから入力された認識対象音声に対して、予め取得されている辞書作成用マイクロホンの周波数特性と前記周波数特性演算手段により求められた入力用マイクロホンの周波数特性との周波数成分の差分のデータを用いて周波数パラメータの補正を行ない、前記辞書作成用マイクロホンの周波数特性に変換する周波数特性補正手段と、

前記周波数特性補正手段により変換された周波数パラメータからなる認識用音声データを認識辞書と照合することにより認識処理し、認識結果を出力する認識処理手段とを具備することを特徴とする音声認識装置。

【請求項3】 複数種類のマイクロホンの周波数特性を予め格納しておくための周波数特性群格納手段と、

前記周波数特性群格納手段に格納されている複数種類のマイクロホンの周波数特性の1つをこれに対応する入力用マイクロホンの周波数特性として選択するためのユーザ指示を受け付けるユーザインタフェース手段と、

前記ユーザインタフェース手段により受け付けられたユーザ指示の示す前記マイクロホンの周波数特性を前記周波数特性群格納手段から選択する周波数特性選択手段と、

前記入力用マイクロホンから入力された認識対象音声に対して、予め取得されている辞書作成用マイクロホンの周波数特性と前記周波数特性選択手段により選択された周波数特性との周波数成分の差分のデータを用いて周波数パラメータの補正を行ない、前記辞書作成用マイクロホンの周波数特性に変換する周波数特性補正手段と、前記周波数特性補正手段により変換された周波数パラメータからなる認識用音声データを認識辞書と照合することにより認識処理し、認識結果を出力する認識処理手段とを具備することを特徴とする音声認識装置。

【請求項4】 異なる周波数特性を有する複数のマイクロホンの周波数特性を辞書作成用マイクロホンの周波数特性に補正するための補正データを予め格納しておくための補正データ群格納手段と、

前記補正データ群格納手段に格納されている複数の補正データの1つをこれに対応する入力用マイクロホンの周波数特性の補正データとして選択するためのユーザ指示を受け付けるユーザインタフェース手段と、

前記ユーザインタフェース手段により受け付けられたユーザ指示の示す前記補正データを前記補正データ群格納手段から選択する補正データ選択手段と、

前記入力用マイクロホンから入力された認識対象音声に対して、前記補正データ選択手段により選択された補正データを用いて周波数パラメータの補正を行ない、前記辞書作成用マイクロホンの周波数特性に変換する周波数特性補正手段と、

前記周波数特性補正手段により変換された周波数パラメータからなる認識用音声データを認識辞書と照合することにより認識処理し、認識結果を出力する認識処理手段とを具備することを特徴とする音声認識装置。

【請求項5】 周波数特性が既知の基準マイクロホンと周波数特性が未知の入力用マイクロホンを用いて、ユーザが発声する音声を前記基準マイクロホンと前記入力用マイクロホンとで同期がとれるように入力し、

前記基準マイクロホン及び前記入力用マイクロホンから入力した音声を周波数分析し、

前記基準マイクロホン及び前記入力用マイクロホンからの入力音声の周波数分析結果の周波数成分の差分のデータを求め、その差分のデータと前記基準マイクロホンの周波数特性とから、前記入力用マイクロホンの周波数特

性を求めておき、

前記入力用マイクロホンから入力された認識対象音声に対して、予め取得されている辞書作成用マイクロホンの周波数特性と前記求めておいた入力用マイクロホンの周波数特性との周波数成分の差分のデータを用いて周波数パラメータの補正を行ない、前記辞書作成用マイクロホンの周波数特性に変換し、認識処理に供するようにしたことを特徴とするマイクロホン周波数特性変換方法。

【請求項6】 予め記憶してある音波形を用いて音を出し、

周波数特性が既知の基準マイクロホンと周波数特性が未知の入力用マイクロホンを用いて、前記出力した音を前記基準マイクロホンと前記入力用マイクロホンとで同期がとれるように入力し、

前記基準マイクロホン及び前記入力用マイクロホンから入力した音を周波数分析し、

前記基準マイクロホン及び前記入力用マイクロホンからの入力音の周波数分析結果の周波数成分の差分のデータを求め、その差分のデータと前記基準マイクロホンの周波数特性とから、前記入力用マイクロホンの周波数特性を求めておき、

前記入力用マイクロホンから入力された認識対象音声に対して、予め取得されている辞書作成用マイクロホンの周波数特性と前記求めておいた入力用マイクロホンの周波数特性との周波数成分の差分のデータを用いて周波数パラメータの補正を行ない、前記辞書作成用マイクロホンの周波数特性に変換し、認識処理に供するようにしたことを特徴とするマイクロホン周波数特性変換方法。

【請求項7】 予め格納しておいた複数種類のマイクロホンの周波数特性の1つをこれに対応する入力用マイクロホンの周波数特性として選択するためのユーザ指示を受け付けて、前記受け付けたユーザ指示の示す前記マイクロホンの周波数特性を選択しておき、

前記入力用マイクロホンから入力された認識対象音声に対して、予め取得されている辞書作成用マイクロホンの周波数特性と前記選択しておいた周波数特性との周波数成分の差分のデータを用いて周波数パラメータの補正を行ない、前記辞書作成用マイクロホンの周波数特性に変換し、認識処理に供するようにしたことを特徴とするマイクロホン周波数特性変換方法。

【請求項8】 異なる周波数特性を有する複数の入力用マイクロホンの周波数特性を辞書作成用マイクロホンの周波数特性に補正するための補正データを予め格納しておき、

前記格納しておいた複数の補正データの1つをこれに対応する入力用マイクロホンの周波数特性の補正データとして選択するためのユーザ指示を受け付けて、前記受け付けたユーザ指示の示す前記補正データを選択しておき、

前記入力用マイクロホンから入力された認識対象音声に

対して、前記選択しておいた補正データを用いて周波数パラメータの補正を行ない、前記辞書作成用マイクロホンの周波数特性に変換し、認識処理に供するようにしたことを特徴とするマイクロホン周波数特性変換方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、マイクロホンの周波数特性の変換機能を持った音声認識装置及びマイクロホン周波数特性変換方法に関する。

10 【0002】

【従来の技術】 一般に音声認識では、同じカテゴリに属する入力音声と認識辞書作成時の音声との類似性を高めるべく、入力音声に対しては様々なパターン変形の要因を排除する対策をとり、認識辞書に対してはできるだけ多くのパターン変形を吸収するような対策がなされる。

【0003】 入力音声に対する対策としては、特に雑音の重畳によるパターン変形を削除するために、雑音の周波数パラメータを推定し、雑音重畳音声の周波数パラメータから雑音成分を引き去るノイズ・サブトラクション法や、マイクロホン特性及び回線特性差を正規化するために回線の周波数特性を2次曲線などで近似し、入力音声に対して周波数特性の補正をする方法、また、電話音声認識では、電話回線歪みの正規化のために、補正用のフィルタにより周波数特性の補正を行ない、スペクトルを平坦化し、電話回線歪みの影響を除去する方法などが提案されている。

【0004】 一方、認識辞書に対する対策としては、雑音の重畳に対して、様々な雑音パターンを重畳した音声パターンを人工的に生成し、その音声を用いて認識辞書を作成する方法、異なるSN比毎に複数の認識辞書を作成しておき、入力音声のSN比を推定して、最も近いSN比の認識辞書を利用する方法、特に認識辞書としてHMM (Hidden Markov Model: 隠れマルコフモデル) を用いる場合には、クリーン音声のHMMと雑音HMMのパラメータから、雑音重畳音声のHMMパラメータを合成する方法などが提案されている。更に、回線歪みに対しても、HMMのパラメータを合成する方法が適用されている。

【0005】 また、音声認識の際の入力音声に対するマイクロホンの周波数特性の補正に関しては、特開平7-84594号に記載の音声認識装置に以下のような提案がある。本例では、認識のために用いる音声入力用のマイクロホンと認識辞書作成のための音声データの収集に用いたマイクロホン(辞書作成用マイクロホン)の2つのマイクロホンを同時に用いて音声入力を行ない、それぞれの音声データを用いて、入力用マイクロホンの特性が辞書作成用マイクロホンの特性になるように適応フィルタの係数を推定し、実際の認識時には、上記の方法で求めたフィルタを用いて、入力音声の補正を行なうというものである。

【0006】

【発明が解決しようとする課題】従来、入力音声に対するマイクロホン特性や回線特性の補正については、事前にそれらの周波数特性を白色雑音などのテスト信号を用いて求めておき、実際の認識の際には、予め作成した周波数特性の補正データを用いて、入力音声の周波数パラメータの変換を行ない、認識処理に供するようにするか、或いは、実際の入力音声から近似手法を用いて周波数特性を推定し、推定した特性を用いて入力音声に対する補正を行なうという方法をとっていた。

【0007】しかし、事前に補正データを求めておく場合には、ユーザが認識の際に利用できるマイクロホンは予め決められていることから、ユーザが使用環境等に応じて任意に選んだマイクロホンでは、適切な周波数特性の補正がなされず、音声認識性能の低下の要因となっていた。

【0008】また、入力音声から周波数特性を推定する場合には、正確な特性を求めることは困難であり、近似的にしか周波数特性の補正ができないため、入力音声に補正を加えても、認識辞書作成に利用したマイクロホン特性（周波数特性）とは異なるため、高い認識性能が得られないという問題があった。

【0009】更に、上述した特開平7-84594号に記載の音声認識装置の例では、ユーザが補正を行なう際に、音声認識辞書作成時のマイクロホンを必ず利用しなければならないが、一般的に認識辞書を作成する際のデータ収集用のマイクロホン（辞書作成用マイクロホン）は、高価であったり、認識語彙セットにより異なるマイクロホンであったりして、ユーザが利用できる環境を整えるためには、ユーザに大きな負担を強いるものとなっていた。

【0010】本発明は、上記の事情を考慮してなされたもので、その目的は、ユーザに音声を発声させる、或いは音を入力することにより、音声入力に利用するマイクロホンと、認識辞書作成に利用したマイクロホンとの周波数特性の差を正確に補正し、高い認識性能を得ることのできる音声認識装置及びマイクロホン周波数特性変換方法を提供することにある。

【0011】本発明の他の目的は、正確な補正データを求める際に、従来、詳細な録音条件を設定する等大きな労力となっていた作業を、認識装置を使うユーザ自身が音入力或いは音声入力を行なうだけで可能とし、作業の効率化を図る音声認識装置及びマイクロホン周波数特性変換方法を提供することにある。

【0012】本発明の更に他の目的は、認識辞書作成用のマイクロホンを必要とせず、且つユーザの労力を大幅に低減できる音声認識装置及びマイクロホン周波数特性変換方法を提供することにある。

【0013】

【課題を解決するための手段】上記の課題を解決するた

めに本発明は、周波数特性が既知のマイクロホン（基準マイクロホン）とユーザが実際に音声認識に利用する周波数特性が未知のマイクロホン（入力用マイクロホン）とを用いて、ユーザの発声した音声の音声データを基準マイクロホンと入力用マイクロホンとで同期がとれるように入力手段（音声入力手段）にてそれぞれ収集し、各データを分析手段（音声分析手段）にて周波数分析し、それぞれの分析結果の差分を周波数特性演算手段にて求めることにより、マイクロホン特性以外の差を相殺し、その差分のデータと基準マイクロホンの周波数特性とから、入力用マイクロホンの周波数特性を求めるようにしたことを特徴とする。

【0014】ところで、音声認識に用いる認識辞書は実際の認識処理以前に予め作成されるものであるため、その辞書作成のための音声データの収集に用いたマイクロホン（辞書作成用マイクロホン）については予め周波数特性を求めておくことが可能である。

【0015】そこで本発明は、上記のように求めた入力用マイクロホンの周波数特性を、辞書作成用マイクロホンの周波数特性に補正するための補正データを、当該辞書作成用マイクロホンの周波数特性と上記入力用マイクロホンの周波数特性との周波数成分の差分をとることで取得し、認識を行なう際には、入力用マイクロホンを介して入力された認識対象音声の周波数パラメータに対して、この補正データを用いて周波数特性補正手段にて周波数特性の変換を行ない、認識処理に供するようにしている。

【0016】このような構成においては、認識対象音声の周波数パラメータを辞書作成用マイクロホンの周波数特性に合わせることができるので、高い認識性能を得ることができる。

【0017】また、従来は正確な補正データを得るためには、詳細な録音条件等を設定してデータ収集を行ない、認識装置とは異なるシステムでオフライン的に処理する必要があり、大きな労力となっていたが、本発明では、ユーザが発声した音声、異なるマイクロホン（基準マイクロホンと入力用マイクロホン）を用いて基準マイクロホンと入力用マイクロホンとで同期がとれるように例えば同時に入力し、それらの周波数分析結果の差分を求めることで、マイクロホン特性以外の差を相殺することができるため、音声認識装置を使うユーザ自身が音声入力を行なうだけで簡単に正確な入力用マイクロホンの周波数特性を取得して正確な補正データを作成することができ、周波数特性測定的大幅な効率化を実現できる。

【0018】ここで、認識辞書と予め求めた辞書作成用マイクロホンの周波数特性を対応付けて格納しておくことにより、上記補正データの作成に当該辞書作成用マイクロホンの周波数特性がそのまま利用できるため、ユーザが補正のたびに辞書作成用マイクロホンを用意する必

要がなく、ユーザの負担を大幅に低減することができる。更に、基準マイクロホンについても、辞書作成用マイクロホンを用いるといった制約は全くなく、この点でもユーザの負担を大幅に低減することができる。

【0019】また、入力用マイクロホンの周波数特性を求める際に、ユーザが音声入力を行なう代わりに、予め記憶してある音波形を用いて音出力手段により音を出力する構成としても構わない。この場合、ユーザは入力用マイクロホンの周波数特性の測定のために音声入力を行なう必要がなく、ユーザの負担を一層軽減できる。

【0020】また、複数種類のマイクロホンの周波数特性を予め格納しておき、ユーザからの指示に応じてユーザが入力用に使用するマイクロホンの周波数特性を選択可能な構成としても構わない。この場合、ユーザが入力用マイクロホンの周波数特性の測定のために音声入力を行わなくて済むだけでなく、入力用マイクロホンの周波数特性を取得するための音声認識装置内での一連の処理が不要となる。

【0021】更に、異なる周波数特性を有する複数のマイクロホンの周波数特性を辞書作成用マイクロホンの周波数特性に補正するための補正データを予め格納しておき、ユーザからの指示に応じてユーザが入力用に使用するマイクロホンの周波数特性の補正データを選択可能な構成としても構わない。この場合、音声認識装置内での補正データの作成処理も不要となる。

【0022】

【発明の実施の形態】以下、本発明の実施の形態につき図面を参照して説明する。

【第1の実施形態】図1は本発明の第1の実施形態に係る音声認識装置の基本構成を示すブロック図である。

【0023】図1の音声認識装置は、主として、音声入力部11、音声分析部12、周波数特性演算部13、基準マイク周波数特性格納部14、入力用マイク周波数特性格納部15、周波数特性補正部16、辞書作成データ収集用マイク周波数特性格納部17、認識処理部18及び認識辞書格納部19から構成される。

【0024】本実施形態において、図1の音声認識装置は、図2に示すような可搬型のパーソナルコンピュータ20を用いて実現される。このパーソナルコンピュータ20の筐体内には内蔵型マイクロホン（以下、内蔵型マイクと略称する）21が予め組み込まれている。また、パーソナルコンピュータ20には、外部からマイク（マイクロホン）が接続可能なマイク入力端子22が設けられている。また、パーソナルコンピュータ20には、ステレオ録音可能な音声入力装置（図示せず）が内蔵されている。

【0025】図1の音声認識装置は、（1）入力用マイクロホンの周波数特性の測定モード、（2）認識モードの2つのモードで作動される。このモードは、ユーザによりキーボード、マウス等の入力手段を操作することで

選択指定できるようになっている。なお、ユーザからのモード指定を受け付けるユーザインタフェースは図1では省略されている。

【0026】ここで、上記2つのモードのうち、本発明の特徴を表す（1）入力用マイクロホンの周波数特性の測定モードでの処理について、図3のフローチャートを参照して説明する。

【0027】入力用マイクロホンの周波数特性の測定モードでは、音声入力部11は、ユーザが発声した音声を、周波数特性が既知の基準マイクロホン（基準マイク）#aと周波数特性が未知の入力用マイクロホン（入力用マイク）#b（いずれも図示せず）とから、マイク# a、# bで同期がとれるように例えば同時に入力する（ステップS1）。音声入力部11は、このステップS1において、各マイク# a、# bから入力した音声を、同じサンプリング周波数でA/D（アナログ/デジタル）変換し、デジタル時系列信号の音声データに変換して時系列順に図示せぬ記憶部に一時記憶する。

【0028】音声入力部11でのA/D変換には、図2のパーソナルコンピュータ20に内蔵のステレオ録音可能な音声入力装置などを利用することが可能である。また、周波数特性が既知のマイク（基準マイク）# aとしては、図2のパーソナルコンピュータ20の筐体内に組み込まれている内蔵型マイク21が利用可能である。なお、図1の音声認識装置を、卓上型のパーソナルコンピュータを用いて実現する場合であれば、例えば当該コンピュータに付属品として添付されるマイクロホンを用いることも可能である。要するに、基準マイク# aは、ユーザが特別に用意しなくても利用可能なマイクロホンで、予め周波数特性が測定されたマイクロホンであればよい。

【0029】一方、周波数特性が未知のマイク（入力用マイク）# bとしては、ユーザが認識に用いるために用意した任意のマイクロホンでよく、例えば図2のパーソナルコンピュータ20のマイク入力端子22に接続して使用すればよい。

【0030】さて音声入力部11は、上記のようにして得られたマイク# a、# bからの音声データに対して、固定時間（例えば、8ミリ秒：以後この単位をフレームと呼ぶ）毎に音声パワーを計算し、パワーの時系列を用いて、発声された音声の始末端の時刻を検出する（ステップS2）。このステップS2において、音声入力部11は、検出した音声の始末端の時刻をもとに、マイク# a、# bからの入力音声の始末端区間内の音声データを抽出し、音声分析部2に渡す。

【0031】音声分析部2は、音声入力部11から渡されたマイク# a、# bそれぞれの始末端区間内の音声データを対象に、例えば高速フーリエ変換（FFT）を用いて周波数分析を行ない、音声データの時系列信号を周波数パラメータの時系列データに変換する（ステップS

3)。

【0032】さて音声分析部12は、上記ステップS3において、マイク#a、#bからの入力音声の周波数パラメータの時系列データを、マイク#a、#b別に全音声区間で平均し、対数変換を行ない、次のようなマイク#a、#bの平均スペクトルのベクトル(平均スペクトルベクトル)A、Bを生成する。但し、マイク#a、#bから例えば単語音声を複数種類入力した場合には、全単語音声について平均を求めるような処理を行なう。

【0033】 $A = (a1, a2, a3, \dots, a256)$: マイク(基準マイク) #aの平均スペクトルベクトル

$B = (b1, b2, b3, \dots, b256)$: マイク(入力用マイク) #bの平均スペクトルベクトル

なお、この平均スペクトルベクトルは、512ポイントのFFTを行なった場合の例であり、1, 2, 3...は各周波数帯域の番号を示す。

【0034】周波数特性演算部13は、音声分析部12により求められたマイク#a、#b毎の平均スペクトルベクトルA、Bと、予め測定されたマイク(基準マイク) #aの周波数特性Faとから、マイク(入力用マイク) #bの周波数特性Fbを次のように求める。ここで、マイク(基準マイク) #aの周波数特性Fa(のデータ)は、基準マイク周波数特性格納部14に予め格納されている。

【0035】まず周波数特性演算部13は、基準マイク#aの平均スペクトルベクトルAと入力用マイク#bの平均スペクトルベクトルBとの差分ベクトルΔBを

$$\Delta B = B - A$$

$$= (b1-a1, b2-a2, b3-a3, \dots, b256-a256)$$

により求める(ステップ4)。このΔBを求めることにより、マイク#a、#bの平均スペクトルベクトルA、Bに含まれる話者特徴、発声内容の音韻特徴、回線歪みなどの共通の特徴が相殺される。

【0036】次に周波数特性演算部13は、基準マイク#aの周波数特性Faを

$$Fa = (fa1, fa2, fa3, \dots, fa256)$$

とすると、この周波数特性Faと上記ΔBとから、入力用マイク#bの周波数特性Fbを

$$Fb = Fa + \Delta B$$

$$= (fa1+(b1-a1), fa2+(b2-a2), fa3+(b3-a3), \dots, fa256+(b256-a256))$$

により求める(ステップ5)。

【0037】周波数特性演算部13は、求めた入力用マイク#bの周波数特性Fb(のデータ)を入力用マイク周波数特性格納部15に格納する(ステップS6)。ここで、マイク#a、#bの平均スペクトルベクトルA、Bに含まれる話者特徴、発声内容の音韻特徴、発声環境などの共通の特徴をXとすると、平均スペクトルベクトルA、Bとマイク#a、#bの周波数特性Fa、Fbとの関係は

$$A = Fa + X$$

$$B = Fb + X$$

と表される。

【0038】これから、

$$\Delta B = B - A$$

$$= (Fb + X) - (Fa + X)$$

$$= Fb - Fa$$

となり、発声内容、話者、発声環境の異なる場合でも共通の特徴Xが相殺されたマイク#a、#bの平均スペクトルベクトルA、Bの差だけが差分ベクトルΔBに表現されていることがわかる。

【0039】以上が、入力用マイクロホンの周波数特性の測定モードでの処理(入力用マイクロホンの周波数特性の推定処理)である。なお、この測定モードにおいてユーザがマイク#a、#bに対して発声する音声(語彙)の内容を、図1の音声認識装置が有するユーザインタフェースによりユーザに対して案内(指示)する構成とすると、ユーザの負担が軽減する。特に、ユーザに案内する音声の内容を、各音韻が同頻度で出現する語彙セットとすると、精度の高い周波数特性の測定が可能となる。

【0040】次に、上記(2)の認識モードでの処理について、図4のフローチャートを参照して説明する。認識モードでは、ユーザが発声した結果、マイク(入力用マイク) #bからの音声が入力部11により取り込まれ、上述したステップS1での処理と同様にA/D変換され、デジタル時系列信号の音声データに変換される(ステップS11)。次に、入力音声の音声データから音声パワーを用いた音声区間(始末端)の検出が行なわれ、その区間内の音声データが抽出される(ステップS12)。

【0041】音声分析部11により抽出された音声区間内の音声データは、音声分析部12にて周波数分析され、周波数パラメータの時系列データに変換される(ステップS13)。この周波数パラメータの時系列データは、音声分析部12から周波数特性補正部16に渡される。

【0042】周波数特性補正部16は、音声分析部12から渡された、入力用マイク#bからの入力音声の周波数パラメータの時系列データを、音声認識辞書作成のための音声データの収集に利用したマイクロホン(以下、辞書作成用マイクと称する)の周波数特性に補正する処理を次のように行なう。

【0043】まず周波数特性補正部16は、上述の入力用マイクの周波数の測定モードで周波数特性格納部15に求められた入力用マイク#bの周波数特性Fbと、予め求められている辞書作成用マイクの周波数特性Ftとを用い、その差分をとることにより、認識時に発声される音声データに対する補正ベクトルCを求める(ステップS14)。

11

【0044】ここで、辞書作成用マイクの周波数特性 F_t は辞書作成データ収集用マイク周波数特性格納部17に予め格納されており、当該周波数特性 F_t を $F_t = (ft1, ft2, ft3, \dots, ft256)$ とすると、補正ベクトル C は $C = F_b - F_t$
 $= (fa1 + (b1 - a1) - ft1, fa2 + (b2 - a2) - ft2, fa3 + (b3 - a3) - ft3, \dots, fa256 + (b256 - a256) - ft256)$ により求められる。

【0045】なお、この補正ベクトル C は、上述した入力用マイクの周波数の測定モードで入力用マイク#bの周波数特性 F_b が求められた段階で算出可能である。したがって、この段階で補正ベクトル C を算出して、例えば入力用マイク周波数特性格納部15（或いは当該周波数特性格納部15に相当する補正データ格納部）に格納しておくならば、認識処理の都度、上記ステップS14のように補正ベクトル C を算出する必要がなくなる。

【0046】さて周波数特性補正部16は、ステップS14により補正ベクトル C を求めると、当該補正ベクトル C を用いて、入力音声の時刻 i 毎に、その時刻の周波数パラメータ S_i に対して、以下の補正を行なうことで、補正後の周波数パラメータ S_i' を得る（ステップS15）。即ち周波数特性補正部16は、周波数パラメータ S_i を

$$S_i = (si1, si2, si3, \dots, si256)$$

とすると、当該周波数パラメータ S_i を補正ベクトル C により

$$S_i' = S_i - C$$

$$= (si1 - fa1 - (b1 - a1) + ft1, si2 - fa2 - (b2 - a2) + ft2, si3 - fa3 - (b3 - a3) + ft3, \dots, si256 - fa256 - (b256 - a256) + ft256)$$

のように補正して、補正後の周波数パラメータ S_i' を得る。

【0047】ここで補正ベクトル C は、マイク#bで入力された音声を辞書作成用マイクで入力した場合の周波数パラメータ S_i' を周波数パラメータ S_i から求めるためのものである。

【0048】即ち、ある時刻 i の周波数パラメータ S_i をマイク#bの周波数特性 F_b と時刻 i の音声に含まれるそれ以外の特徴 Y とで表すと、 $S_i = F_b + Y$ となる。

【0049】また補正ベクトル C は、前記したように $C = F_b - F_t$ ある。

【0050】以上から、周波数パラメータの推定値 S_i' は、

$$S_i' = S_i - C$$

$$= (F_b + Y) - (F_b - F_t)$$

$$= F_t + Y$$

となり、 S_i' が辞書作成用マイクの周波数特性 F_t と

12

時刻 i のそれ以外の音声特徴 Y からなることがわかる。

【0051】次に周波数特性補正部16は、時刻 i 毎の補正後の周波数パラメータ S_i' を認識処理で用いる音声特徴ベクトルに変換する（ステップS16）。このステップS16の詳細を、音声特徴ベクトルとしてバンドパスフィルタ出力を用いる場合を例に述べる。

【0052】まず周波数特性補正部16は、周波数特性の補正を行なった周波数パラメータ S_i' に対し、指数変換を行ない対数スケールから線形スケールに変換する。次に周波数特性補正部16は、周波数パラメータ S_i' のベクトルを低次元の音声特徴パラメータベクトルに次元圧縮する。ここでは、例えばBarkスケールなどを用い、音声特徴ベクトルの各チャンネルに相当する複数の周波数パラメータのチャンネルの平均を求め、対数変換することにより、目的とする低次元の音声特徴ベクトルを求める。

【0053】こうして得られた音声特徴ベクトルは、認識処理部18において、認識辞書格納部19に格納されている音声認識辞書との照合がなされ（ステップS17）、認識語彙の中で最も高い尤度をもつ語彙が認識結果として例えばディスプレイ（図示せず）に出力される（ステップS18）。認識辞書としては、例えば同じカテゴリに属する複数のデータから作成した平均ベクトルパターンや、統計的手法としてHMMなどを用いることができる。また、認識処理部18での照合の際には、一般に行なわれるDP（Dynamic Programming：動的計画法）などを利用することができる。

【0054】以上に述べた図1の構成の音声認識装置で適用した処理、即ち入力用マイクロホンの周波数特性の測定モードでの（音声入力部11、音声分析部12及び周波数特性演算部13の）処理と、認識モードでの（音声入力部11、音声分析部12、周波数特性補正部16及び認識処理部18の）処理は、その処理を図2のパーソナルコンピュータ20等により実行させるためのプログラムを記録した記憶媒体、例えばフロッピーディスク23をパーソナルコンピュータ20に装着して、当該フロッピーディスク23に記録されているプログラムをパーソナルコンピュータ20で読み取り実行させることにより実現される。

【0055】なお、前記第1の実施形態では、ユーザは、入力用マイクロホンの周波数特性の測定モードにおいて、測定に必要な種々の単語音声等を発声する必要がある。そこで以下では、上記の測定モードでの単語音声等の発声を不要にした第2の実施形態につき説明する。

【第2の実施形態】図5は本発明の第2の実施形態に係る音声認識装置の基本構成を示すブロック図であり、図1と同一部分には同一符号を付してある。

【0056】図5の音声認識装置が図1の音声認識装置と異なる点は、入力用マイクロホンの周波数特性の測定モードにおいて、測定のためにユーザが音声を発声する

代わりに、予め用意された音信号が利用可能な構成となっていることである。そのため図5の構成では、図1の構成とは異なって、音出力のためのユーザからの指示入力を受け付けるユーザインタフェース51と、このユーザインタフェース51により音出力のためのユーザ指示入力を受け付けられた場合に、音出力を行なう音出力部52と、この音出力部52による音出力に必要な音波形（のデータ）を予め記憶しておくための出力音記憶部53とが新たに設けられている。なお、ユーザインタフェース51は、ユーザからのモード指定等の受け付けも行なう。

【0057】図5の音声認識装置において、ユーザが入力用マイクロホンの周波数特性の測定モードを指定し、更に音出力を指示すると、ユーザインタフェース51は音出力部52を起動する。すると音出力部52は、出力音記憶部53に予め記憶されている音波形を読み込み、その音波形に従う音出力を行なう。

【0058】ここで、出力音記憶部53に記憶されている音波形は雑音であったり、人が発声した音声でもよい。但し、マイクロホンの周波数特性を測定するためにあらゆる周波数成分が含まれていることが望ましいため、白色雑音を記憶しておき出力することが望ましい。また、記憶した波形が音声である場合には、数多くの種類の音韻が含まれる言葉を数多く記憶しておき出力することが望ましい。

【0059】音声入力部11は、周波数特性が既知のマイクロホンであるマイク（基準マイク）#aと、周波数特性が未知のマイクロホンであるマイク（入力用マイク）#bの両方を用いて、音出力部52から出力された音の取り込みをそれぞれ同時に行なう。このとき、各マイク# a、# bから入力される音は、同じサンプリング周波数でA/D変換され、ディジタル時系列信号の音声データに変換される。この音声入力部11での動作は、入力の対象が、前記第1の実施形態のようなユーザが発声した音声ではなくて、音出力部52からの出力音となった点を除けば、前記第1の実施形態におけるステップS1と同様であり、上記A/D変換には、図2のパーソナルコンピュータ20に内蔵のステレオ録音可能な音声入力装置などを利用することが可能である。

【0060】また、基準マイク# aとしては、図2のパーソナルコンピュータ20の筐体内に組み込まれている内蔵型マイク21が利用可能である。なお、図1の音声認識装置を、卓上型のパーソナルコンピュータを用いて実現する場合であれば、例えば当該コンピュータに付属品として添付されるマイクロホンを用いることも可能である。要するに、基準マイク# aは、ユーザが特別に用意しなくても利用可能なマイクロホンで、予め周波数特性が測定されたマイクロホンであればよい。

【0061】一方、入力用マイク# bとしては、ユーザが認識に用いるために用意した任意のマイクロホンでよ

く、例えば図2のパーソナルコンピュータ20のマイク入力端子22に接続して使用すればよい。

【0062】音声入力部11における以降の処理、即ち、マイク# a、# bからの入力音の始末端区間内の音声データを抽出する処理も、前記第1の実施形態（におけるステップS2）と同様である。更に、入力用マイク# bの周波数特性Fbを求めるための以降の処理（音声分析部12及び周波数特性演算部13での処理）も、前記第1の実施形態（におけるステップS3～S6）と同様である。この他、認識モードでの（音声入力部11、音声分析部12、周波数特性補正部16及び認識処理部18の）処理も、前記第1の実施形態（におけるステップS11～S18）と同様である。

【0063】以上に述べた第1及び第2の実施形態では、入力用マイクロホンの周波数特性を音声認識装置内で測定することを前提としており、そのための周波数特性測定モードにおいて、測定に必要な音または音声をマイク# a、# bにより同時に入力し、その入力音声をもとにマイク（入力用マイク）# bの周波数特性Fbを求める必要があった。そこで以下では、ユーザが使用する入力用マイクロホンによっては、周波数特性測定モードを不要とすることができるようにした第3の実施形態につき説明する。

【第3の実施形態】図6は本発明の第3の実施形態に係る音声認識装置の基本構成を示すブロック図であり、図1と同一部分には同一符号を付してある。

【0064】図6の音声認識装置が図1の音声認識装置と異なる点は、ユーザによるマイク特性（入力用マイクロホンの周波数特性）の選択指定を可能とするためのユーザインタフェース61と、そのためのマイク特性選択メニュー画面等の表示に用いられるディスプレイ62と、予め求められた種々のマイクロホンの周波数特性が格納されているマイク周波数特性群格納部63と、ユーザが選択指定した入力用マイクの周波数特性をマイク周波数特性群格納部63から選択して（周波数特性演算部13によって求められた場合と同様に）入力用マイク周波数特性格納部15に格納する入力用マイク周波数特性選択部64とが設けられていることである。但し、図1の構成の音声認識装置においても、機能は異なるもののユーザインタフェース61に相当するユーザインタフェースが存在し、またマイク特性選択メニュー画面は表示されないもののディスプレイ62に相当するディスプレイが存在する。

【0065】ここで、図6の音声認識装置の主としてユーザインタフェース61が提供するユーザによるマイク特性の選択指定のためのインタフェース機能について説明する。

【0066】まず本実施形態では、ユーザインタフェース61により、図7に示すようなマイク特性選択メニュー画面がディスプレイ62に表示される。図7のメニ

一画面の例では、予め求められた4つのマイク（マイクロホン）と「その他」の計5つの選択肢（メニュー項目）が表示されている。

【0067】ユーザは、自身が入力用に利用するマイク（入力用マイク#b）の製造会社名、型番などがわかっており、それに合致するものがメニュー画面上にある場合には、該当するマイク（のメニュー項目）をマウス等の入力手段で選択する。

【0068】ユーザインタフェース61は、ユーザによるメニュー画面からのマイク選択のための指示入力を受け付けると、入力用マイク周波数特性選択部64を起動する。すると入力用周波数特性選択部64は、ユーザが選択指定したマイクの周波数特性をマイク周波数特性群格納部63から選択的に取り出して（入力用マイク#bの周波数特性Fbとして）入力用マイク周波数特性格納部15に格納する。

【0069】このように本実施形態においては、ユーザが入力用に利用するマイク（図2のパーソナルコンピュータ20により音声認識装置が実現される場合であれば、マイク入力端子22に接続して使用するマイク）のメニュー項目が図7のマイク特性選択メニュー画面に存在するならば、そのメニュー項目をユーザが選択指定するだけで、マイク周波数特性測定（推定）のための音声入力をすることなく、図3のフローチャートに示したような入力用マイクの周波数特性測定モード時処理を行なうことなく、指定の入力用マイク（#b）の周波数特性を入力用マイク周波数特性格納部15内に得ることができる。

【0070】この入力用マイク周波数特性格納部15内に得られたユーザ指定の入力用マイク（#b）の周波数特性は、（前記第1の実施形態における入力用マイクの周波数特性測定モード時の処理で得られた入力用マイク#bの周波数特性と同様に）認識モード時の入力用マイク#bからの入力音声に対する補正に利用される。

【0071】一方、図7のマイク特性選択メニュー画面で示されている製造会社名や型番（の情報）では入力用に利用するマイク（入力用マイク#b）のメニュー項目を判断できない場合には、ユーザは画面右下の「詳細」ボタンをクリックすればよい。この場合、ユーザインタフェース61は、例えばマイクの外形や寸法、周波数特性図などの詳細情報をディスプレイ62に表示し、ユーザが選択するための詳細な情報を提供する。

【0072】更に、これらの詳細情報でも判断できない場合には、入力用に利用するマイク（入力用マイク#b）のメニュー項目は存在しないものとして、ユーザはメニュー項目「その他」を選択する。

【0073】ユーザインタフェース61は、図7のマイク特性選択メニュー画面上で「その他」が選択された場合、前記第1の実施形態で述べた入力用マイクの周波数特性の測定モードを自動設定する。この場合、ユーザイ

ンタフェース61はまず、図8（a）に示すような入力用マイクの周波数特性の測定を開始する旨のガイダンス画面をディスプレイ62に表示する。

【0074】ここでユーザが「次へ」ボタンをクリックすると、ユーザインタフェース61は測定するマイクを使用する位置にセットするようユーザに促す図8（b）に示すような画面に切り換える。このとき図6の音声認識装置が図2の可搬型（ノート型）のパーソナルコンピュータ20を用いて実現されており、基準マイク#aとして表示モニタ周辺の筐体内に内蔵されているマイクを用いるよう設定されているとすると、「モニタに対して正面の姿勢を取ってください」などのメッセージを表示する。一方、マイクが添付された卓上型のパーソナルコンピュータを用いて音声認識装置が実現されており、添付されたマイクを基準マイクとするよう設定されているとすると、「添付されたマイクをセットして下さい」などの表示を行なうなどして、基準マイクのセットを促す。

【0075】図8（a）の画面でユーザが「次へ」ボタンをクリックすると、ユーザインタフェース61はユーザに発声させる語彙を表示してユーザの発声を促す図8（c）に示すような画面に切り換える。この画面には、ユーザに発声させる語彙の表示フィールド81と、発声単語数の表示フィールド82が設けられている。

【0076】ここで、ユーザに発声させる語彙セットは、できるだけ様々な周波数帯の成分が含まれる語彙セットにすることが望ましい。これは、各周波数帯の特性を正確に測定するためである。このような語彙セットを予め複数セット用意しておき、入力用マイクの周波数特性の測定を複数回行なう場合には、語彙セットを変えて、前回の測定の際に発声した音声データも利用し、マイク特性の補正を行なうよう制御するなどして、できるだけ多くの音声データを利用し、各周波数帯の成分が均等になるようにするとよい。また、各周波数帯の成分が均等になるような語彙セットとしては、例えば音韻の出現頻度が同頻度になるような単語セット（音韻バランス単語セット）などを用いるのが現実的である。

【0077】ユーザは、図8（c）の画面の案内に従って、フィールド81に表示された語彙を発声する。すると、図6の音声認識装置内では、前記第1の実施形態における入力用マイクの周波数特性の測定モード時と同様の処理が行なわれる。但し、全ての語彙の発声が完了するまでは、図3のステップS3における平均スペクトルベクトルA、Bを求める処理以降は待たされる。この場合、（ステップS3の前半部において）マイク#a、bからの入力音声の始末端区間内の音声データから音声分析部12により変換される周波数パラメータの時系列データは記憶部に一時記憶される。

【0078】さて、ユーザインタフェース61（システム）が提示した全ての語彙についての発声が終了する

と、或いはユーザが必要と考える個数の語彙についての発声が終了して「次へ」ボタンをクリックすると、ユーザインタフェース61は、図9(a)に示すようなマイク特性(入力マイクの周波数特性)の計算を行なう旨の表示を行なう。

【0079】同時にユーザインタフェース61は、音声分析部12に対して平均スペクトルベクトルA、Bの計算開始を指示する。これを受けて音声分析部12は、記憶部に一時記憶されている周波数パラメータの時系列データ、即ちユーザが発声した全ての語彙についてのマイク# a、bからの入力音声の周波数パラメータの時系列データから、マイク# a、# bの平均スペクトルのベクトル(平均スペクトルベクトル)A、Bを生成する。これ以降、入力用マイク# bの周波数特性を求めて周波数特性格納部15に格納するまでの処理(図3のステップS14~S16に相当する周波数特性演算部13の処理)は、前記第1の実施形態と同様である。

【0080】ユーザインタフェース61は、以上の処理の進捗状況を検出し、その進捗状況を図9(a)の表示画面上でスライドバー等によりフィールド91に表示する。ユーザインタフェース61は、入力用マイク# bの周波数特性(マイク特性)の測定が終了したならば、求めたマイク特性を保存するか否かの指示をユーザに促す図9(b)に示すような画面に切り換える。

【0081】ここでユーザが保存するを選択した場合には、ユーザインタフェース61は、保存するファイル名の入力フィールド92と、ID(識別名)の入力フィールド93を持つ画面を表示して、ユーザにファイル名とIDを入力させる。するとユーザインタフェース61は、入力用マイク周波数特性選択部64を制御して、周波数特性演算部13により入力用マイク周波数特性格納部15に得られている入力用マイク# bの周波数特性を、ユーザ指定のファイル名とIDに対応付けてマイク周波数特性群格納部63に保存させる。このユーザ指定のIDは次回からのマイク特性選択メニュー画面(図7参照)のメニューに追加表示される。これによりユーザは、今回入力用に使用したマイクを再度入力用に使用したい場合には、当該マイクに対応するIDをマイク特性選択メニュー画面上で選択するだけで、当該マイクの周波数特性を選択することができる。

【0082】最後にユーザインタフェース61は、図9(c)の画面に切り換えることで、マイク特性選択に関する一連の処理が終了したことをユーザに伝える。さて、以上のようにして、ユーザが選択指定した入力用マイク# bの周波数特性、または入力用マイクの周波数特性の測定モードで取得された入力用マイク# bの周波数特性が入力用マイク周波数特性格納部15に格納されると、以降の認識モードでユーザが発声した音声、即ち入力用マイク# bからの入力音声の周波数パラメータは、当該周波数特性格納部15に格納された周波数特性を利

用じて(周波数特性補正部16により図4のステップS15、S16と同様の処理で)辞書作成用マイクの周波数特性に適応するように補正される。

【0083】認識処理部18は、周波数特性補正部16による周波数特性補正後の入力音声の音声特徴ベクトルと認識辞書格納部19内の音声認識辞書との照合を行ない、認識語彙の中で最も高い尤度をもつ語彙を認識結果として出力する。

【0084】認識処理部18から出力される認識結果は、図示せぬ経路を介してユーザインタフェース61に渡される。ユーザインタフェース61は、この認識結果をディスプレイ62に表示して、ユーザに当該認識結果の正誤を判定させる。そしてユーザインタフェース61は、認識結果が誤っていると判定された場合、現在の入力用マイク# bの周波数特性をそのまま採用し、単にユーザが再度の発声を行なうか、マイク周波数特性群格納部63から別の周波数特性を選択するか、或いは入力用マイク# bの周波数特性を改めて測定して入力用マイク周波数特性格納部15の内容を更新するかを、ディスプレイ62を通してユーザに選択指定させる。

【0085】以上に述べた第3の実施形態では、複数種類のマイクロホンの周波数特性を予め周波数特性群格納部63に用意しておいて、その中からユーザ指定のマイクロホンの周波数特性を選択して入力用マイク周波数特性格納部15に格納するものとして説明した。この場合、前記第1の実施形態の認識モードでの処理(ステップS14)で述べたように、周波数特性格納部15内の入力用マイク# bの周波数特性Fbと、辞書作成データ収集用周波数特性格納部17内の辞書作成用マイクの周波数特性Ftとを用い、その差分をとることにより、認識用に発声される音声データに対する補正ベクトルCを求める必要がある。

【0086】そこで、この補正ベクトルCを、複数種類のマイクロホン別に、そのマイクロホンの周波数特性と、辞書作成用マイクの周波数特性Ftとから予め算出して、マイク周波数特性群格納部63(または当該周波数特性群格納部63に相当する補正データ群格納部)に格納しておき、その中からユーザ指定のマイクロホンの周波数特性に対応する補正ベクトルCを入力用マイク周波数特性選択部64(或いは周波数特性選択部64に相当する補正データ選択部)により選択して、入力用マイク周波数特性格納部15(或いは当該周波数特性格納部15に相当する補正データ格納部)に格納する構成とするならば、認識モードでの周波数特性補正部16による周波数パラメータの補正処理において、当該補正ベクトルCがそのまま利用できるため、周波数特性補正部16による補正ベクトルCの算出処理を不要とすることができる。

【0087】なお、以上の実施形態で述べた周波数パラメータなどの分析条件は、高速フーリエ変換に限るもの

ではなく、例えばLPC (Linear Predictive Coding) 分析を用いてもよいし、フィルタ分析を用いてもよい。

【0088】また、図7乃至図9に示した表示画面は、一例であって、この内容に限定されるものではない。この他、本発明は前記実施形態に限定されるものではなく、その要旨を逸脱しない範囲で、種々変形して実施することができる。

【0089】

【発明の効果】以上詳記したように本発明によれば、周波数特性が既知のマイクロホン（基準マイクロホン）とユーザが実際に音声認識に利用する周波数特性が未知のマイクロホン（入力用マイクロホン）の両方を用いて、音声認識装置を使うユーザ自身が音声入力を行なうだけ、或いは音声認識装置に対して音出力を指示するだけで、入力用マイクロホンの周波数特性を簡単に且つ正確に求めることができ、認識性能の高度化および周波数特性測定的大幅な効率化を実現できる。

【0090】また本発明によれば、予め求めた辞書作成マイクロホンの周波数特性を利用可能な構成とすることで、補正のたびに辞書作成マイクロホンを用意する必要がなく、ユーザの負担を大幅に低減することができる。

【図面の簡単な説明】

【図1】本発明の第1の実施形態に係る音声認識装置の基本構成を示すブロック図。

【図2】図1の音声認識装置を実現するパーソナルコンピュータの外観を示す図。

【図3】図1の音声認識装置における入力用マイクロホンの周波数特性の測定モードでの処理を説明するためのフローチャート。

【図4】図1の音声認識装置における認識モードでの処理を説明するためのフローチャート。

【図5】本発明の第2の実施形態に係る音声認識装置の基本構成を示すブロック図。

【図6】本発明の第3の実施形態に係る音声認識装置の基本構成を示すブロック図。

【図7】図6の音声認識装置で適用されるマイク特性選択メニュー画面の一例を示す図。

【図8】図7のマイク特性選択メニュー画面上で「その他」が選択された結果、入力用マイクの周波数特性の測定モードが設定された場合の、ユーザに対するガイダンス画面の画面遷移例の一部を示す図。

【図9】図7のマイク特性選択メニュー画面上で「その他」が選択された結果、入力用マイクの周波数特性の測定モードが設定された場合の、ユーザに対するガイダンス画面の画面遷移例の残りを示す図。

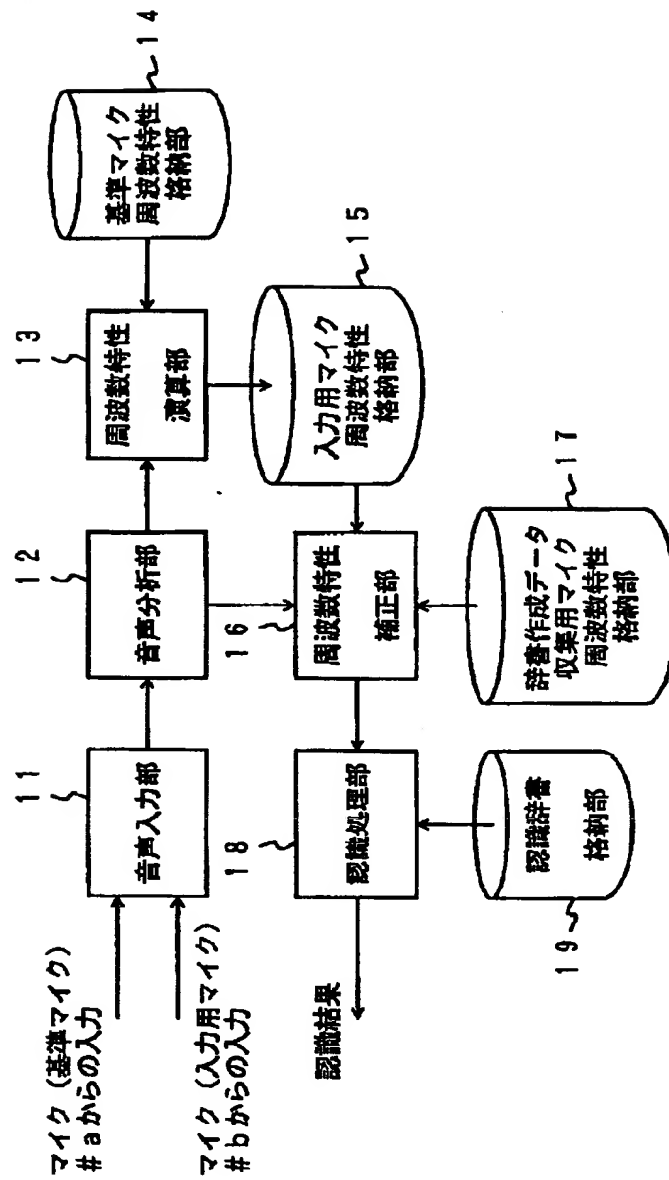
【符号の説明】

- 11…音声入力部、
- 12…音声分析部、
- 13…周波数特性演算部、
- 14…基準マイク周波数特性格納部、
- 15…入力用マイク周波数特性格納部、
- 16…周波数特性補正部、
- 17…辞書作成データ収集用マイク周波数特性格納部、
- 18…認識処理部、
- 19…認識辞書格納部、
- 51, 61…ユーザインタフェース、
- 52…音出力部、
- 53…出力音記憶部、
- 62…ディスプレイ、
- 63…マイク周波数特性群格納部、
- 64…入力用マイク周波数特性選択部。

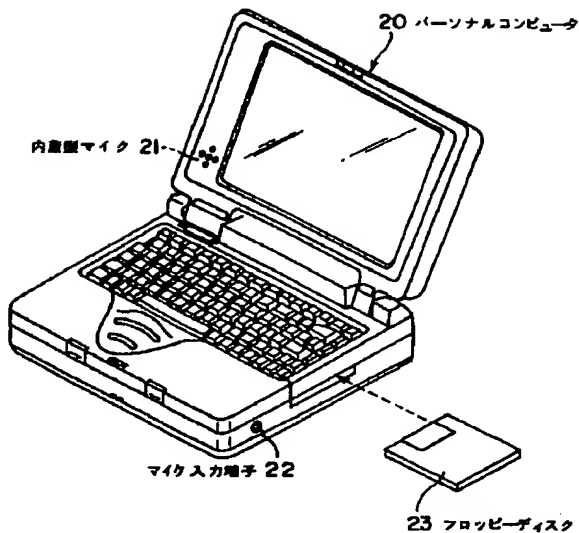
【図7】

- ◎ A社 単一指向性マイク (型番A1, A2)
- B社 無指向性マイク (型番B1, B2, B3)
- C社 単一指向性マイク (型番C1, C4)
- C社 無指向性マイク (型番C2, C3)
- その他

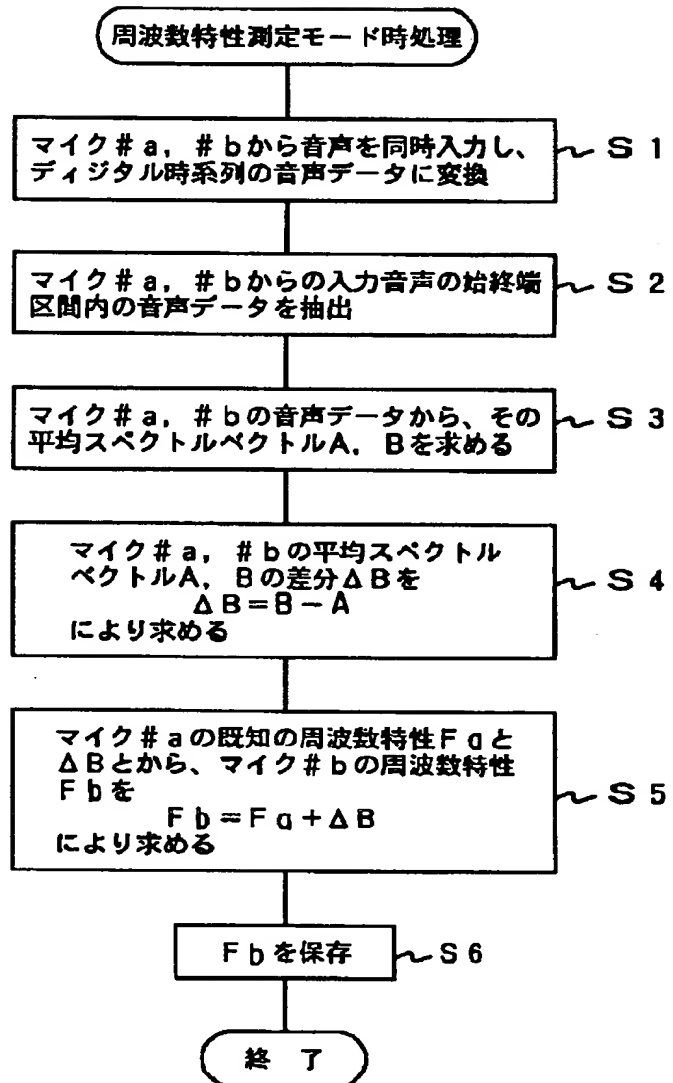
【図1】



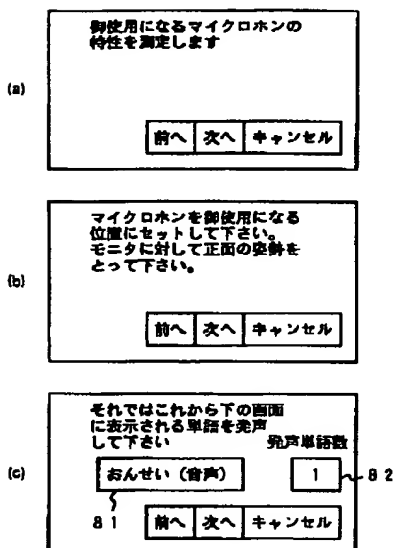
【図2】



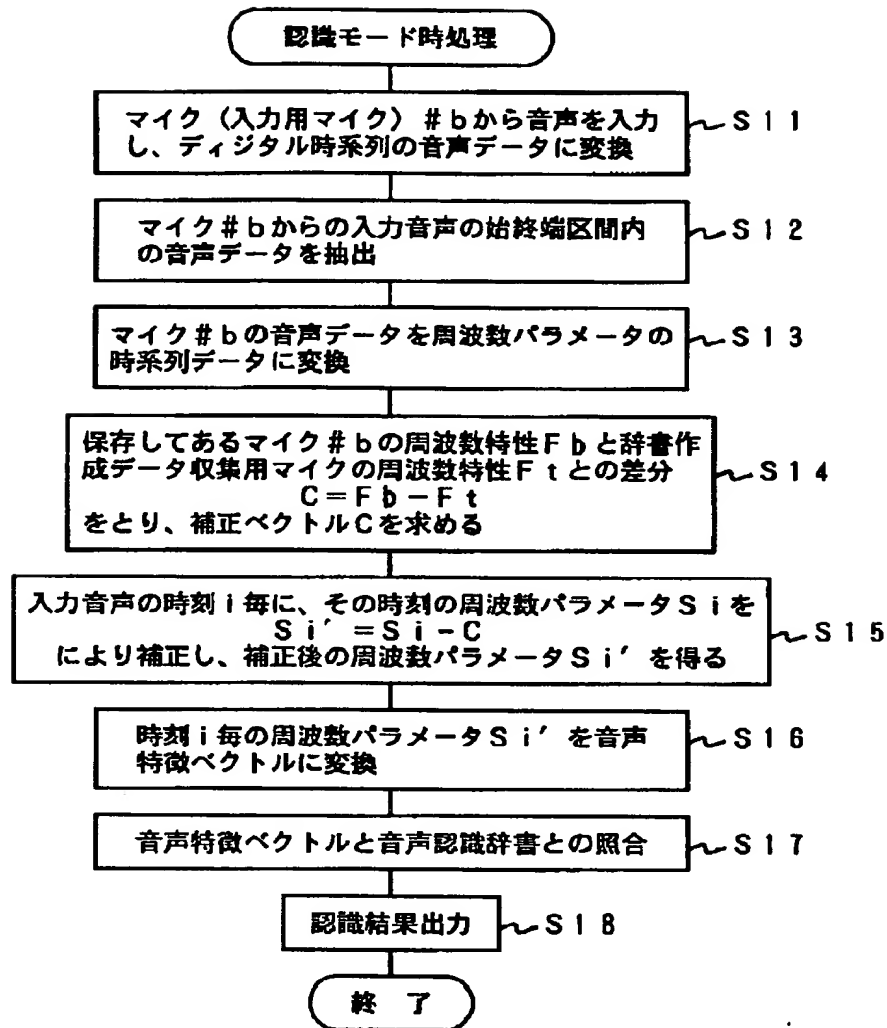
【図3】



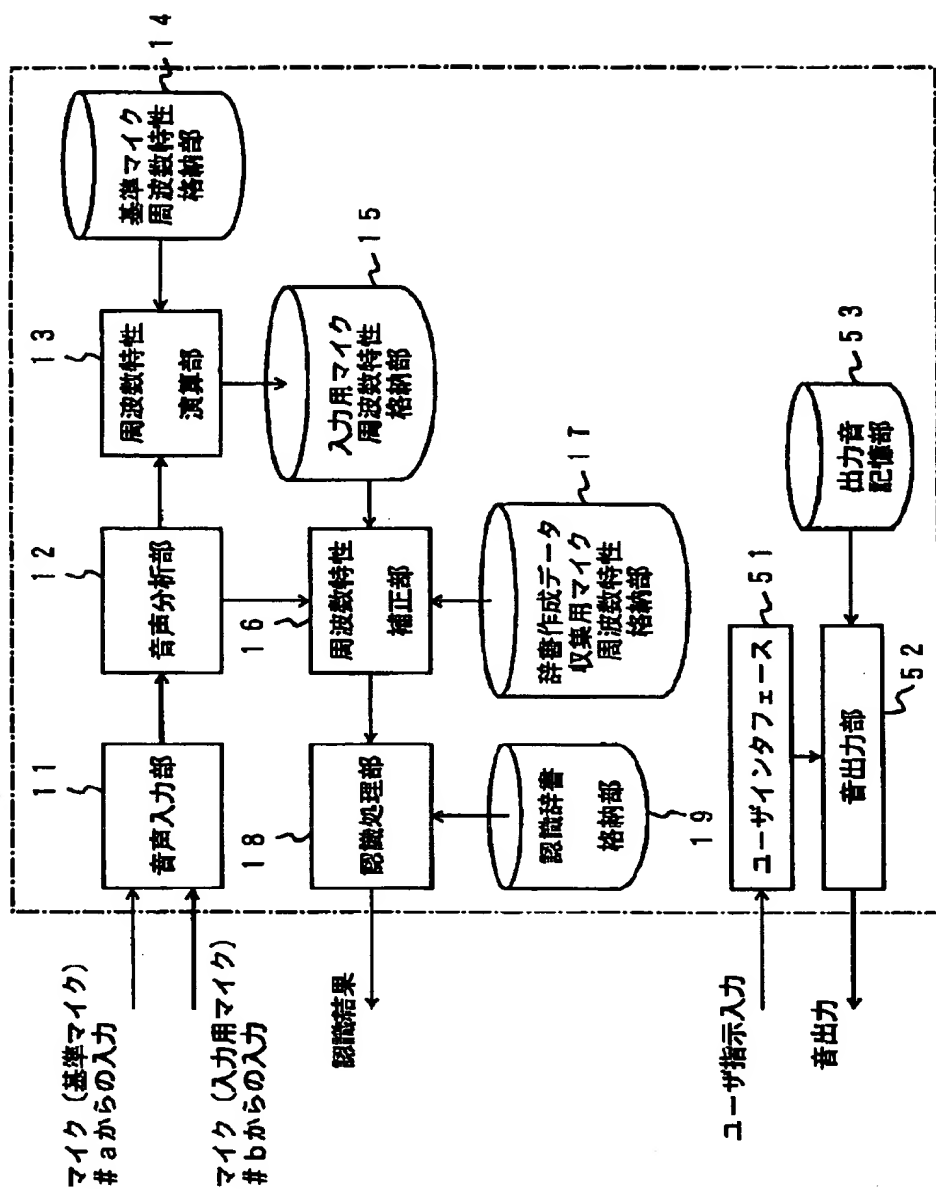
【図8】



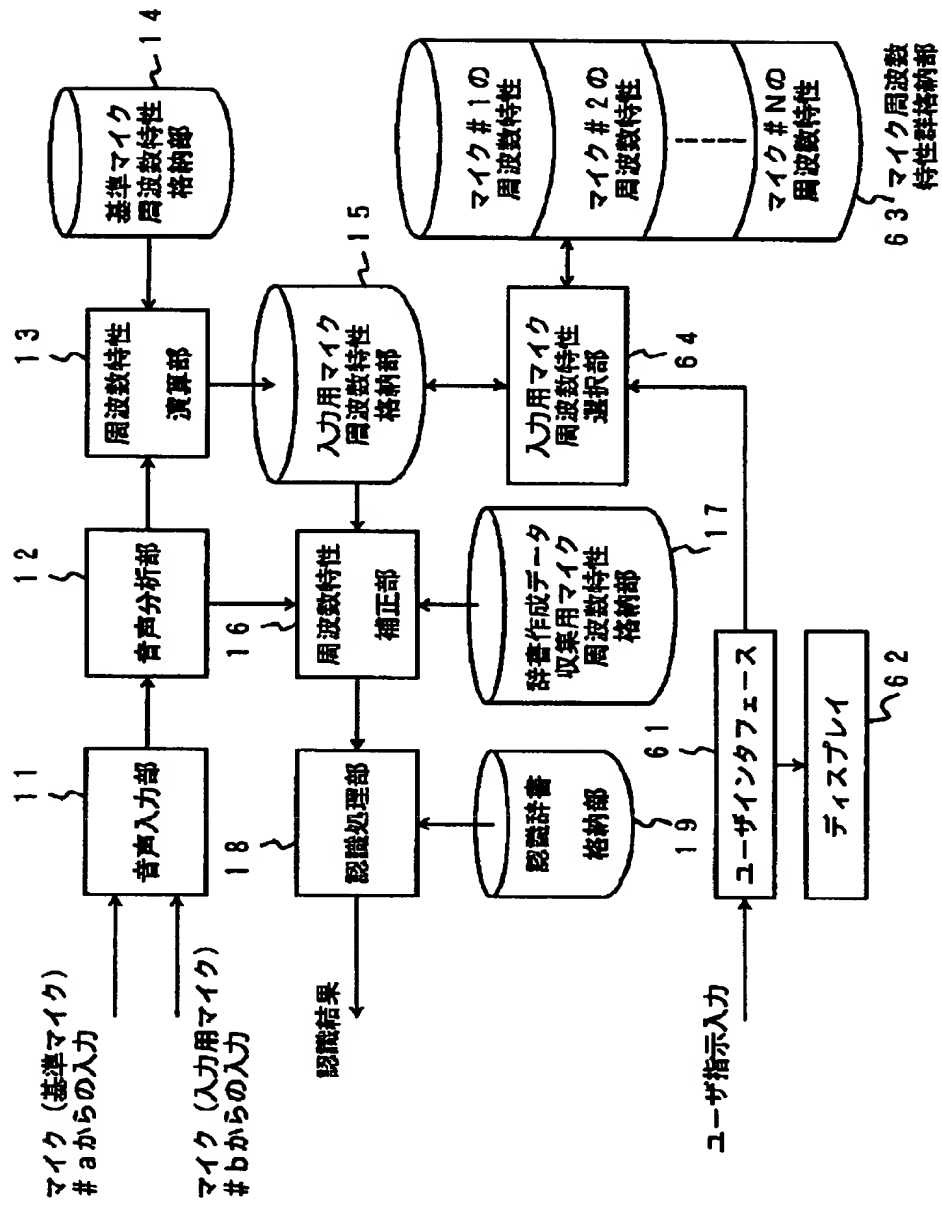
【図4】



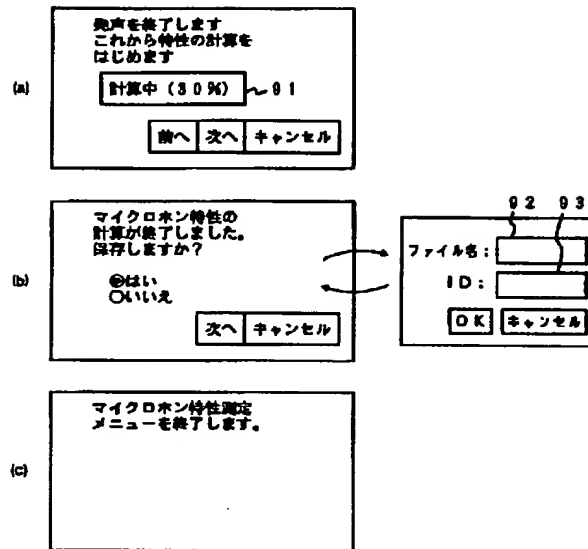
【図5】



【図6】



【図9】



フロントページの続き

(72)発明者 坪井 宏之
大阪府大阪市北区大淀中1丁目1番30号
株式会社東芝関西支社内